

Some Emerging Design Criteria for Interim Assessments within a Balanced Assessment System

Nathan Dadey

Center for Assessment

Reidy Interactive Lecture Series

Portsmouth, NH

September 27, 2018



This work is licensed under
a Creative Commons
Attribution 4.0
International License.

Context

Some states have:

- increasingly taken interest in interim assessments¹, perhaps in response to their continued widespread use, as well as federal attention.
- developed programs aimed at providing interim assessments or improving practice interim.
 - Generally targeted to the district level, although schools could leverage many of the same resources.

¹As defined by Perie et al. (2009): “assessments administered during instruction to evaluate students’ knowledge and skills relative to a specific set of academic goals in order to inform policymaker or educator decisions at the classroom, school, or district level. The specific interim assessment designs are driven by the purposes and intended uses, but the results of any interim assessment must be reported in a manner allowing *aggregation* across students, occasions, or concepts” (2009, p. 6, emphasis added).

Top-Down & Bottom-Up

- From a systems perspective, such designs can be called “top-down” in the sense that the state is attempting to improve the top levels of the system – the district and state.
- An alternative, “bottom-up” design is one in which the district attempts to improve the classroom and district levels, with hopefully some connection to the state level. (see Conley, 2018 for some examples)
 - Haven’t seen examples of systems that develop from the classroom-level on up.

Systems Perspective

- Considering how these efforts function from a system of assessment perspective leads us to ask questions like:
 - What is the relationship between the state- and district-level assessments?
 - How are the assessments complementing one another in terms of coherence, comprehensiveness and continuousness (plus utility & efficiency)? In addition, how tightly the levels coupled?

Additional Design Considerations

Addressing these types of questions leads to two additional, overlapping criteria that could be helpful:

- **Learning Targets:** what are the depth and breath of the interim assessments, in relation to the state-level assessments (and vice-versa)? (Gong, 2010)
- **Degree of Modularity:** if there are multiple assessments:
 - At what grain size is the content targeted to?
 - How flexible is the timing of administration?

¹In this presentation, the terms interim assessments and district-level assessment are interchangeable.

Learning Targets

- Breath
 - Full coverage vs. selected subset of the domain
 - If divided up among multiple assessments, how is the content divided?
- Depth
 - Full range of depth vs. focus on a specific level of depth (perhaps for a specific subset of the domain)

Degree of Modularity

Fixed Designs

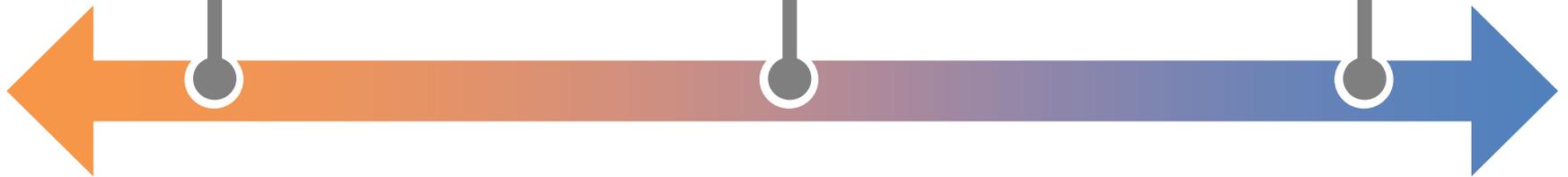
A single assessment that measures the entire domain

Block Designs

Multiple assessments, each measuring a chunk of the domain

Modular Designs

Multiple assessments, each measuring a very small chunk of the domain



Fixed

Modular

Fixed Designs

Fixed Designs

A single assessment that measures the entire domain

- Broadly measure the domain (sometimes in ways that mirror the state-level accountability assessment)
- Administered as needed or within specified administration windows

Block Designs

Block Designs

Multiple
assessments, each
measuring a chunk
of the domain

- Measures a predefined set of the domain (e.g., content domain, unit grouping)
- Administered on demand, sometimes within pre-specified windows
- Examples:
 - Delaware's end of unit assessments
 - Louisiana's EAG ELA assessments
 - Wyoming's interim assessments
 - Smarter Balanced Interim Blocks

Modular Designs

Modular Designs

Multiple assessments, each measuring a very small chunk of the domain

- Measures a fine grained part of the domain
- Administered on demand, sometimes within pre-specified window, and sometimes with recommended groupings
- Examples:
 - Wyoming's modular assessments
 - Kentucky's through course tasks

Continued Challenges

- Flexibility
- Scoring & Scaling
- Funding
- Control
- Communication
- Examples

Questions

- If these categories are useful, how do we help those engaging in district & state design work choose between different designs and balance the pros and cons of each design?
 - When would multiple designs be advocated (e.g., Wyoming's suite of interim (in this case a fixed design) and modular assessments?)

Some Example Designs

- A tightly coupled design: Delaware's NGSS Assessment System
 - Interim: Block design aligned to unit content
 - State Summative: A fixed design aligned to a subset of the domain, but with a focus on transfer
- A more loosely coupled design: Kentucky's NGSS Assessment System
 - Interim: Modular design (task bank) with tasks aligned to individual standards, likely at a deep level of complexity, with reporting requirements
 - State Summative: A fixed design broadly aligned to the standards